ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Applications of Random Set Theory in Econometrics

Ilya Molchanov[1] and Francesca Molinari[2]

[1]Department of Mathematical Statistics and Actuarial Science, University of Bern, 3012 Bern, Switzerland

[2]Department of Economics, Cornell University, Ithaca, New York 14850; email: fm72@cornell.edu

Annu. Rev. Econ. 2014. 6:229–51

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi: 10.1146/annurev-economics-080213-041205

JEL codes: C1, C2, C3, C57

Keywords

capacity functional, Aumann expectation, support function, partial identification

Abstract

In recent years, the econometrics literature has shown a growing interest in the study of partially identified models, in which the object of economic and statistical interest is a set rather than a point. The characterization of this set and the development of consistent estimators and inference procedures for it with desirable properties are the main goals of partial identification analysis. This review introduces the fundamental tools of the theory of random sets, which brings together elements of topology, convex geometry, and probability theory to develop a coherent mathematical framework to analyze random elements whose realizations are sets. It then elucidates how these tools have been fruitfully applied in econometrics to reach the goals of partial identification analysis.

1. INTRODUCTION

Random set theory is concerned with the development of a coherent mathematical framework to study random objects whose realizations are sets. Such objects appeared a long time ago in statistics and econometrics in the form of confidence regions, which can be naturally described as random sets. The first idea of a general random set in the form of a region that depends on chance appears in Kolmogorov (1950), originally published in 1933. A systematic development of the theory of random sets did not occur until later, stimulated by the study in general equilibrium theory and decision theory of correspondences and nonadditive functionals, as well as the need in image analysis, microscopy, and materials science of statistical techniques to develop models for random sets, estimate their parameters, filter noisy images, and classify biological images.

These and other related applications of set-valued random variables induced the development of statistical models for random sets, furthered the understanding of their distributions, and led to the seminal contributions of Choquet (1953/1954), Aumann (1965), and Debreu (1967) and to the first self-contained treatment of the theory of random sets given by Matheron (1975). Since then, the theory has expanded in several directions, developing its relationship with convex geometry, various limit theorems for random sets, set-valued processes, etc. An account of the modern theory of random sets is available in Molchanov (2005).

More recently, the development within econometrics of partial identification analysis has provided a new and natural area of application for random set theory. Partially identified econometric models appear when the available data and maintained assumptions do not suffice to uniquely identify the statistical functional of interest, might this be finite or infinite dimensional, even as data accumulate (see Tamer 2010 for a review and Manski 2003 for a systematic treatment). For this class of models, partial identification proposes that econometric analysis should study the set of values for the statistical functional that are observationally equivalent, given the available data and credible maintained assumptions; in this article, this set of values is referred to as the functional's sharp identification region. The goals of the analysis are to obtain a tractable characterization of the sharp identification region, to provide methods for estimating it, and to conduct tests of hypotheses and make confidence statements about it.

Conceptually, partial identification predicates a shift of focus from single-valued to set-valued objects, which renders it naturally suited for the use of random set theory as a mathematical framework to conduct identification analysis and statistical inference, to unify a number of special results, and to produce novel general results. The random set approach complements the more traditional one, based on mathematical tools for (single-valued) random vectors, that proved extremely productive since the beginning of the research program in partial identification (see, e.g., Manski 1995 for results on identification and Horowitz & Manski 2000, Imbens & Manski 2004, Chernozhukov et al. 2007, and Andrews & Soares 2010 for results on statistical inference).

A lack of point identification can generally be traced back to a collection of random variables that are consistent with the available data and maintained assumptions. In many cases, this collection of observationally equivalent random variables is equal to the family of selections of a properly specified random closed set, and random set theory can be applied to describe their distribution and to derive statistical properties of estimators that rely on them. Specific examples discussed in detail in this article include interval data and finite static games with multiple equilibria. In the first case, the random variables consistent with the data are those that lie in the interval with probability one. In the second case, the random variables consistent with the modeling assumptions are the ones that represent equilibria of the game.

To fruitfully apply random set theory for identification and inference, econometricians need to carry out three fundamental steps. First, they need to define the random closed set that is relevant to the problem under consideration using all information given by the available data and maintained assumptions. This is a delicate task but is one that is typically carried out in identification analysis regardless of whether random set theory is applied. Second, they need to determine how the observable random variables relate to this random closed set. Often, one of two cases occurs: Either the observable variables determine a random set to which the (unobservable) variable of interest belongs with probability one (e.g., the interval data example), or the (expectation of the) (un)observable variable belongs to (the expectation of) a random set determined by the model (e.g., the games with multiple equilibria example). Finally, they need to determine which tool from random set theory should be utilized. To date, new applications of random set theory to econometrics have fruitfully exploited Aumann expectations and their support functions, (Choquet) capacity functionals, and laws of large numbers and central limit theorems for random sets.

In this article, we begin in Section 2 by reviewing these basic elements of random set theory. Then in Section 3 we review the econometrics literature that has applied them for identification analysis. Econometrics applications to statistical inference are discussed in Section 4. Section 5 concludes.

The goal of this review is to provide a guide to the study of random set theory using comprehensive textbooks such as Molchanov (2005), from the perspective of applications in econometrics (this goal is further developed in Molchanov & Molinari 2014). Our view is that the instruction of random set theory could be fruitfully incorporated into PhD-level field courses in econometrics on partial identification and in microeconomics on decision theory. Important prerequisites for the study of random set theory include measure theory and probability theory; good knowledge of convex analysis and topology is beneficial but not essential.

2. RANDOM SET THEORY REVIEW

Throughout this article, we use capital Latin letters to denote sets and random sets. We use lowercase Latin letters for random vectors. We denote parameter vectors and sets of parameter vectors by θ and Θ, respectively. We let $(\Omega, \mathfrak{F}, \mathbf{P})$ denote a nonatomic probability space on which all random variables and random sets are defined, where Ω is the space of elementary events equipped with σ-algebra $\mathfrak{F}$ and probability measure $\mathbf{P}$. We denote the Euclidean space by $\mathbb{R}^d$ and equip it with the Euclidean norm (which is denoted by $\|\cdot\|$).

The theory of random closed sets generally applies to the space of closed subsets of a locally compact Hausdorff second countable topological space $\mathbb{K}$ (see Molchanov 2005). Unless otherwise specified, in this article we let $\mathbb{K} = \mathbb{R}^d$ to simplify the exposition. Denote by $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{K}$ the collection of closed, open, and compact subsets of $\mathbb{R}^d$, respectively. Let $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ and $\mathbb{B}^d = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$ denote the unit sphere and unit ball in $\mathbb{R}^d$, respectively. Given a set $A \subset \mathbb{R}^d$, let $\mathrm{conv}(A)$ denote its convex hull.

2.1. Random Sets

The conventional theory of random sets deals with random closed sets. An advantage of this approach is that random points (i.e., random sets that are singletons) are closed, so the theory of random closed sets includes the classical case of random points or random vectors as a special case. A random closed set is a measurable map $X : \Omega \mapsto \mathcal{F}$, where measurability is defined by specifying the family of functionals of X that are random variables. In specifying this family, a balance is sought between a need for weak conditions, so that there is a large class of examples of random sets, and a need for strict conditions, so that important functionals of random sets are random variables. This trade-off results in the following definition.

Definition 1: A map X from a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ to $\mathcal{F}$ is called a random closed set if

$$X^{-}(K) = \{\omega : X(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}$$

for each compact set $K \subset \mathbb{R}^d$.

In other words, a random closed set is a measurable map from the given probability space to the family of closed sets equipped with the σ-algebra generated by the families of closed sets $\{F \in \mathcal{F} : F \cap K \neq \emptyset\}$ for all $K \in \mathcal{K}$. A random compact set is defined as a random closed set that is compact with probability one so that almost all values of X are compact sets. A convex random set is defined similarly so that $X(\omega)$ is a convex closed set for almost all ω.

Example 1 (interval data): Interval data is a commonplace problem in economics and the social sciences. Let $Y = [y_L, y_U]$ be a random interval on $\mathbb{R}$ where y_L and y_U are two (dependent) random variables such that $y_L \leq y_U$ almost surely. If $K = [a, b]$, then

$$\{Y \cap K \neq \emptyset\} = \{y_L < a, y_U \geq a\} \cup \{y_L \in [a, b]\} \in \mathfrak{F}$$

because y_L and y_U are random variables. Measurability for all compact sets $K \subset \mathbb{R}$ follows from similar arguments.

Example 2 (entry game): Consider a two-player entry game as in Tamer (2003), where each player j can choose to enter ($y_j = 1$) or to stay out of the market ($y_j = 0$). Let ε_1 and ε_2 be two random variables, and $\theta_1 \leq 0$ and $\theta_2 \leq 0$ be two parameters. Let players' payoffs be $\pi_j = y_j(\theta_j y_{3-j} + \varepsilon_j)$, $j = 1, 2$. Each player enters the game if and only if $\pi_j \geq 0$. Then, for given values of θ_1 and θ_2, the set of pure-strategy Nash equilibria, denoted Y_θ, is depicted in **Figure 1** as a function of ε_1 and ε_2. The figure shows that for $(\varepsilon_1, \varepsilon_2) \notin [0, -\theta_1) \times [0, -\theta_2)$, the equilibrium of the game is unique, whereas for $(\varepsilon_1, \varepsilon_2) \in [0, -\theta_1) \times [0, -\theta_2)$, the game admits multiple equilibria and the corresponding realization of Y_θ has cardinality two. An equilibrium is guaranteed to exist because we assume $\theta_1 \leq 0, \theta_2 \leq 0$. To see that Y_θ is a random closed set, notice that in this example, one can take $\mathbb{K} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ and that all its subsets are compact. Then

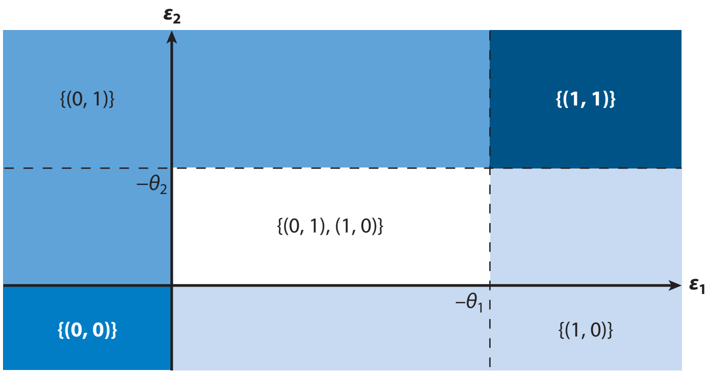

Figure 1

The set of pure-strategy Nash equilibria of a two-player entry game as a function of ε_1 and ε_2.

$$\{Y_\theta \cap K \neq \emptyset\} = \{(\varepsilon_1, \varepsilon_2) \in G_K\} \in \mathfrak{F},$$

where G_K is a Borel set determined by the chosen K. For example, if $K = \{(0, 0)\}$, then $G_K = (-\infty, 0) \times (-\infty, 0)$. Measurability follows because ε_1 and ε_2 are random variables.

2.2. Capacity Functional and Containment Functional

Definition 1 means that X is explored by its hitting events (i.e., the events where X hits a compact set K). The corresponding hitting probabilities have an important role in the theory of random sets; hence, we define them formally here, together with a closely related functional.

Definition 2: (*a*) A functional $T_X(K) : \mathcal{K} \mapsto [0, 1]$ given by

$$T_X(K) = \mathbf{P}\{X \cap K \neq \emptyset\}, \quad K \in \mathcal{K},$$

is called the capacity (or hitting) functional of X.

(*b*) A functional $C_X(F) : \mathcal{F} \mapsto [0, 1]$ given by

$$C_X(F) = \mathbf{P}\{X \subset F\}, \quad F \in \mathcal{F},$$

is called the containment functional of X. We write $T(K)$ and $C(F)$ instead of $T_X(K)$ and $C_X(F)$ in cases in which no ambiguity occurs.

In random set theory, the capacity functional is important because it uniquely determines the probability distribution of a random closed set X (see Molchanov 2005, chapter 1, section 1.2). We note that the containment functional defined on the family of all closed sets F yields the capacity functional extended to open sets $G = F^c$ as

$$T(G) = \mathbf{P}\{X \cap G \neq \emptyset\} = 1 - \mathbf{P}\{X \subset G^c\} = 1 - C(F)$$

and then by approximation determines T on all compact sets. Therefore, the containment functional defined on the family of closed sets also determines the distribution of X. If X is a random compact set, the containment functional defined on the family of compact sets suffices to determine the distribution of X. If $X = \{\xi\}$ is a random singleton with distribution P_ξ, then $T_X(K) = \mathbf{P}\{\xi \in K\} = P_\xi(K)$ and $C_X(F) = \mathbf{P}\{\xi \in F\} = P_\xi(F)$; in other words, T_X and C_X coincide and become the probability distribution of ξ. In particular, then T_X and C_X are additive so that $T_X(K_1 \cup K_2) = T_X(K_1) + T_X(K_2)$ for disjoint K_1 and K_2, and similarly for C_X. In general, however, T_X is a subadditive functional, and C_X is a superadditive functional. This is because in situations in which X contains more than a single point with positive probability, it might hit two disjoint sets simultaneously, so that $T_X(K_1 \cup K_2) \leq T_X(K_1) + T_X(K_2)$, and it might be a subset of a union of sets, but of neither of them alone, so that $C_X(F_1 \cup F_2) \geq C_X(F_1) + C_X(F_2)$.

Example 3 (interval data): Consider again the random interval $Y = [y_L, y_U]$. Then $T_Y(\{a\}) = \mathbf{P}\{y_L \leq a \leq y_U\}$ and $T_Y([a, b]) = \mathbf{P}\{y_L < a, y_U \geq a\} + \mathbf{P}\{y_L \in [a, b]\}$. Similarly, $C_Y([a, b]) = \mathbf{P}\{y_L \geq a, y_U \leq b\}$.

Example 4 (entry game): Consider the setup in Example 2. Then for $K = \{(0, 1)\}$ we have $T(\{0, 1\}) = \mathbf{P}\{\varepsilon_1 < -\theta_1, \varepsilon_2 \geq 0\}$ and $C(\{0, 1\}) = \mathbf{P}\{\varepsilon_1 < -\theta_1, \varepsilon_2 \geq 0\} - \mathbf{P}\{0 \leq \varepsilon_1 < -\theta_1, 0 \leq \varepsilon_2 < -\theta_2\}$. For $K = \{(1, 0), (0, 1)\}$ we have $T(\{(1, 0), (0, 1)\}) = C(\{(1, 0), (0, 1)\}) = 1 - \mathbf{P}\{\varepsilon_1 \geq -\theta_1, \varepsilon_2 \geq -\theta_2\} - \mathbf{P}\{\varepsilon_1 < 0, \varepsilon_2 < 0\}$. One can similarly obtain $T(K)$ and $C(K)$ for each $K \subset \mathbb{K}$.

2.3. Selections and Artstein's Inequalities

Ever since the seminal work of Aumann (1965), it has been common to think of random sets as bundles of random variables—the selections of the random sets.

Definition 3: For any random set X, a (measurable) selection of X is a random vector x such that $x(\omega) \in X(\omega)$ almost surely. We denote by $\mathrm{sel}(X)$ the set of all selections from X.

We often call x a measurable selection to emphasize that x is measurable itself. Recall that a random closed set is defined on the probability space $(\Omega, \mathfrak{F}, \mathbf{P})$ and, unless stated otherwise, almost surely means $\mathbf{P}$ almost surely. A possibly empty random set clearly does not have a selection, so unless stated otherwise, we assume that all random sets are almost surely nonempty, which in turn guarantees the existence of measurable selections (Molchanov 2005, theorem 1.2.13). One can view selections as curves taking values in the tube that is the graph of the random set X.

Example 5 (interval data): Consider again the random interval $Y = [y_L, y_U]$. Then $\mathrm{sel}(Y)$ is the family of all $\mathfrak{F}$-measurable random variables y such that $y(\omega) \in [y_L(\omega), y_U(\omega)]$ almost surely.

To tie the notion of selections to the more traditional approaches in econometrics, we note that each selection of Y can be represented as follows. Take a random variable r such that $\mathbf{P}\{0 \le r \le 1\} = 1$; r's distribution is left unspecified and can be any probability distribution on [0, 1]. Let

$$y = r y_L + (1 - r) y_U.$$

Then $y \in \mathrm{sel}(Y)$. Tamer (2010) gives this representation of the random variables in the interval $[y_L, y_U]$.

Example 6 (entry game): Consider the set Y_θ plotted in **Figure 1**. Let $\Omega^M = \{\omega \in \Omega : \varepsilon(\omega) \in [0, -\theta_1) \times [0, -\theta_2)\}$. Then for $\omega \notin \Omega^M$, the set Y_θ has only one selection because the equilibrium is unique. For $\omega \in \Omega^M$, Y_θ contains a rich set of selections, which can be obtained as

$$y(\omega) = \begin{cases} (0,1) & \text{if } \omega \in \Omega_1, \\ (1,0) & \text{if } \omega \in \Omega_2, \end{cases}$$

for all measurable partitions $\Omega_1 \cup \Omega_2 = \Omega^M$.

Artstein (1983) and Norberg (1992) provide a necessary and sufficient condition that relates the distribution of the selections of the random set X to the capacity (and containment) functional of X. This is considered a fundamental result in random set theory because it allows one to characterize the distribution of bundles of random vectors that constitute random sets.

Theorem 1 (Artstein): A probability distribution μ on $\mathbb{R}^d$ is the distribution of a selection of a random closed set X in $\mathbb{R}^d$ if and only if

$$\mu(K) \le T(K) = \mathbf{P}\{X \cap K \neq \emptyset\} \tag{1}$$

for all compact sets $K \subset \mathbb{R}^d$. Equivalently, μ on $\mathbb{R}^d$ is the distribution of a selection of a random closed set X in $\mathbb{R}^d$ if and only if

$$\mu(F) \ge C(F) = \mathbf{P}\{X \subset F\} \tag{2}$$

for all closed sets $F \subset \mathbb{R}^d$. If X is a compact random closed set, it suffices to check Equation 2 for compact sets F only.

A proof is provided in Molchanov (2005, corollary 1.4.44) and Molchanov & Molinari (2014).

Importantly, we note that if μ from Theorem 1 is the distribution of some random vector x, then it is not guaranteed that $x \in X$ almost surely (e.g., x can be independent of X). Theorem 1 means that for each such μ, it is possible to construct a random variable x that has distribution μ and that belongs to X almost surely. In other words, one couples x and X on the same probability space. Hence, the nature of the domination condition in Equation 1 can be traced to the ordering, or first-order stochastic dominance, concept for random variables. Two random variables x and y are stochastically ordered if $F_x(t) \geq F_y(t)$ (i.e., $\mathbf{P}\{x \leq t\} \geq \mathbf{P}\{y \leq t\}$ for all t). In this case, it is possible to find two random variables x' and y' distributed as x and y, respectively, such that $x' \leq y'$ almost surely. A standard way to determine these random variables is to set $x' = F_x^{-1}(u)$ and $y' = F_y^{-1}(u)$ by applying the inverse cumulative distribution functions to the same uniformly distributed random variable u. One then speaks about the ordered coupling of x and y. We note that the stochastic dominance condition can also be written as $\mathbf{P}\{x \in A\} \leq \mathbf{P}\{y \in A\}$ for $A = [t, \infty)$ and all $t \in \mathbb{R}$. Such a set A is increasing (or upper) (i.e., $x \in A$ and $x \leq y$ implies $y \in A$). Using the probabilities of upper sets, this domination condition can be extended to any partially ordered space. In particular, this leads to the condition for the ordered coupling for random closed sets Z and X obtained by Norberg (1992). Two random closed sets Z and X can be realized on the same probability space as random sets Z' and X' having the same distribution as Z and X, respectively, and so that $Z' \subset X'$ almost surely, if and only if the probabilities that Z has a nonempty intersection with any finite family of compact sets $K_1, \ldots, K_n$ are dominated by those of X. If the set Z is a singleton (e.g., $Z = \{x\}$), such condition can be substantially simplified and reduces to the one in the inequalities in Equation 1.

If Equation 1 holds, then μ is called selectionable. In this article, we refer to Equation 1 as Artstein's inequalities. It follows immediately that T_X and C_X equal the upper envelope and the lower envelope, respectively, of all probability measures that are dominated by T_X and dominate C_X. Specifically, given

$$\mathbb{P}_X = \left\{\mu : \mu(K) \leq T_X(K) \forall K \in \mathcal{K}\right\} = \left\{\mu : \mu(F) \geq C_X(F) \forall F \in \mathcal{F}\right\},$$

we have (see Molchanov 2005, theorem 1.5.13)

$$\begin{aligned} T_X(K) &= \sup\left\{\mu(K) : \mu \in \mathbb{P}_X\right\}, \quad K \in \mathcal{K}, \\ C_X(F) &= \inf\left\{\mu(F) : \mu \in \mathbb{P}_X\right\}, \quad F \in \mathcal{F}. \end{aligned}$$

Because of this, the functionals T_X and C_X are also called coherent upper and lower probabilities. In general, the upper and lower probabilities are defined as envelopes of families of probability measures that do not necessarily stem from a random closed set.

2.4. Aumann Expectations and Support Functions

The space of closed sets is not linear, which causes substantial difficulties in defining the expectation of a random set. One approach, inspired by Aumann (1965) and pioneered by Artstein & Vitale (1975), relies on representing a random set using the family of its selections and considering the set formed by their expectations.

If X possesses at least one integrable selection, then X is called integrable. In this case, only existence is important; for example, X being a segment on the line with one end point equal to zero and the other one equal to a Cauchy distributed random variable is integrable because it possesses a selection that equals zero almost surely, regardless of the fact that its other end point is not integrable. The family of all integrable selections of X is denoted by $\mathrm{sel}^1(X)$.

Definition 4: The (selection or Aumann) expectation of an integrable random closed set X is given by

$$\mathbf{E}X = \mathrm{cl}\left\{\int_{\Omega} x d\mathbf{P} : x \in \mathrm{sel}^1(X)\right\}.$$

If X is almost surely nonempty and its norm $\|X\| = \sup\{\|x\| : x \in X\}$ is an integrable random variable, then X is said to be integrably bounded, and all its selections are integrable. In this case, the family of expectations of these integrable selections is already closed, and there is no need to take an additional closure as required in Definition 4 (see Molchanov 2005, theorem 2.1.24).

The selection expectation depends on the probability space used to define X (see Molchanov 2005, section 2.1.2). In particular, if the probability space is nonatomic and X is integrably bounded, the selection expectation $\mathbf{E}X$ is a convex set regardless of whether X might be nonconvex itself (Molchanov 2005, theorem 2.1.15). This convexification property of the selection expectation implies that the expectation of the closed convex hull of X equals the closed convex hull of $\mathbf{E}X$, which in turn equals $\mathbf{E}X$. It is then natural to describe the Aumann expectation through its support function because this function traces out a convex set's boundary; therefore, knowing the support function is equivalent to knowing the set itself (see **Figure 2** and Equation 3 below).

Definition 5: Let K be a convex set. The support function of K is

$$h_K(u) = \sup\{\langle k, u\rangle : k \in K\}, \quad u \in \mathbb{R}^d,$$

where $\langle k, u\rangle$ denotes the scalar product.

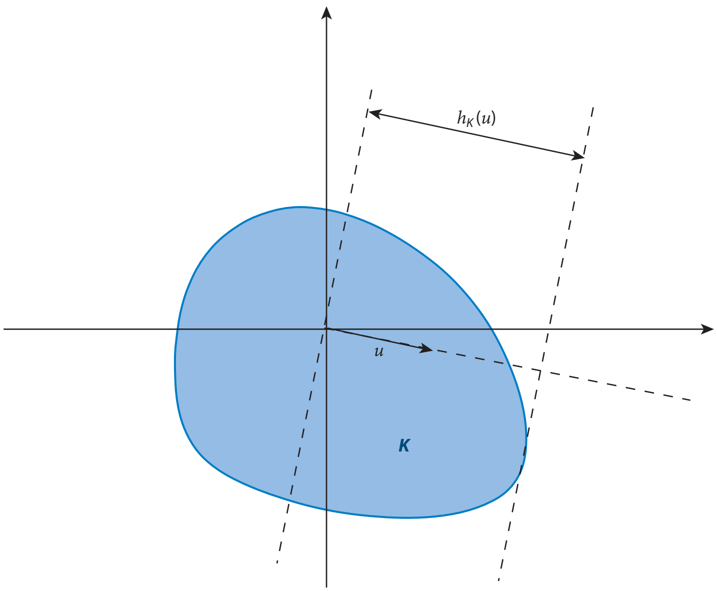

Figure 2

The support function of K in direction u is the signed distance of the support plane to K with exterior normal vector u from the origin; the distance is negative if and only if u points into the open half space containing the origin (Schneider 1993, p. 37).

Note that the support function is finite for all u if K is bounded and is sublinear (positively homogeneous and subadditive) in u. Hence, it can be considered only for $u \in \mathbb{B}^d$ or $u \in \mathbb{S}^{d-1}$. Moreover, one has

$$K = \cap_{u \in \mathbb{B}^d} \{k : \langle k, u \rangle \leq h_K(u)\} = \cap_{u \in \mathbb{S}^{d-1}} \{k : \langle k, u \rangle \leq h_K(u)\}. \tag{3}$$

The great advantage of working with the support function of the Aumann expectation stems from the following result.

Theorem 2: If an integrably bounded random set X is defined on a nonatomic probability space, or if X is almost surely convex, then

$$\mathbf{E} h_X(u) = h_{\mathbf{E}X}(u), \quad u \in \mathbb{R}^d.$$

A proof is provided in Molchanov (2005, theorem 2.1.22).

This implies that one does not need to calculate the Aumann expectation directly by looking at all selections but can simply work with the expectation of the support function of the random set.

2.5. Limit Theorems for Sums of Random Sets

Consider a sequence of independently and identically distributed (i.i.d.) random sets X_i, $i = 1, \ldots, n$, where the notion of i.i.d. in this case corresponds to the requirements that

$$\mathbf{P}\{X_1 \cap K_1 \neq \emptyset, \ldots, X_n \cap K_n \neq \emptyset\} = \prod_{i=1,\ldots,n} \mathbf{P}\{X_i \cap K_i \neq \emptyset\} \quad \forall K_1, \ldots, K_n \in \mathcal{K},$$
$$\mathbf{P}\{X_i \cap K \neq \emptyset\} = \mathbf{P}\{X_j \cap K \neq \emptyset\} \quad \forall i, j, \forall K \in \mathcal{K}.$$

Random set theory provides laws of large numbers and central limit theorems for Minkowski sums of i.i.d. random sets that mimic the familiar ones for random vectors. Given two sets A and B in $\mathbb{R}^d$, and scalars λ and γ in $\mathbb{R}$, define the dilated set $\lambda A = \{r \in \mathbb{R}^d : r = \lambda a,\ a \in A\}$ and let the Minkowski sum of the sets λA and γB be defined as $\lambda A + \gamma B = \{r \in \mathbb{R}^d : r = \lambda a + \gamma b,\ a \in A, b \in B\}$. The Minkowski summation is a commutative and associative operation. Notably, however, it is not an invertible operation: Given two sets A and B, it might be impossible to find a set C such that $A + C = B$ (e.g., this happens if A is a ball and B is a rectangle). Hence, whereas with random variables one expresses limit theorems by taking the difference between a sample average of the variables and their expectation (and normalizing it with a growing sequence), in the case of random sets, one considers the (normalized) Hausdorff distance between the Minkowski average of the sets and their Aumann expectation, where the Hausdorff distance between two sets A and B is defined as

$$\begin{aligned} \rho_H(A, B) &= \inf\{r > 0 : A \subset B^r, B \subset A^r\} \\ &= \max\{\mathbf{d}_H(A, B), \mathbf{d}_H(B, A)\}, \end{aligned}$$

where $A^r = \{a : \mathbf{d}(a, A) \leq r\}$ denotes the r-envelope of A, and

$$\mathbf{d}_H(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

denotes the directed Hausdorff distance from A to B.

The limit theorems rely on three key steps. First, attention is restricted to convex random sets, and the sets are represented as elements of a functional space by means of their support function. This is useful because the sum of the support functions of a sequence of sets is equal to the support function of the Minkowski sum of the sets:

$$h_{\frac{1}{n}\sum_{i=1}^{n} X_i}(u) = \frac{1}{n}\sum_{i=1}^{n} h_{X_i}(u).$$

Second, an embedding theorem given by Hörmander (1954) guarantees that the space of compact and convex subsets of $\mathbb{R}^d$ endowed with the Hausdorff metric can be isometrically embedded into a closed convex cone in the space of continuous functions on the unit sphere endowed with the uniform metric so that

$$\rho_H(X_1, X_2) = \sup_{u\in\mathbb{S}^{d-1}} \left\| h_{X_1}(u) - h_{X_2}(u) \right\|.$$

Finally, the Shapley-Folkman-Starr theorem states that for $K_1, \ldots, K_n$ being any subsets of $\mathbb{R}^d$,

$$\rho_H\Big(K_1 + \cdots + K_n, \mathrm{conv}(K_1 + \cdots + K_n)\Big) \le \sqrt{d} \max_{1\le i\le n} \|K_i\|.$$

Because for a sequence of i.i.d. integrably bounded random sets $X_1, \ldots, X_n$ we have that $n^{-1} \max\|X_i\|$ converges to zero almost surely, taking a Minkowski average yields asymptotic convexification. Hence, the Hausdorff distance between the Minkowski average of not necessarily convex, but integrably bounded sets and the Minkowski average of their convex hulls converges to zero.

Putting these steps together, one obtains

$$\begin{aligned} \left| \rho_H\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \mathbf{E}X\right) - \rho_H\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{conv}(X_i), \mathbf{E}X\right) \right| &\le \rho_H\left(\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} \mathrm{conv}(X_i)\right) \\ &= \mathcal{O}_p\left(\frac{1}{n}\right). \end{aligned}$$

Hence, one obtains Theorem 3 below by using a law of large numbers and a central limit theorem for continuous-valued random variables (the i.i.d. average of support functions minus their expectation), together with Hörmander's embedding theorem, which converts it into a result for the Hausdorff distance between Minkowski averages of convex random sets and their Aumann expectation, and the Shapley-Folkman-Starr theorem, which allows one to lift the requirement of convexity of the sets.

Theorem 3 (law of large numbers for random sets): Let $X, X_1, X_2, \ldots$ be i.i.d. integrably bounded random closed sets in $\mathbb{R}^d$. Then

$$\rho_H\left(\frac{X_1 + \cdots + X_n}{n}, \mathbf{E}X\right) \to 0 \quad \text{almost surely as } n \to \infty.$$

A proof is given in Molchanov (2005, theorem 3.1.6).

Theorem 4 (central limit theorem for random sets): Let $X, X_1, X_2, \ldots$ be i.i.d. random closed sets in $\mathbb{R}^d$ such that $\mathbf{E}\|X\|^2 < \infty$. Then

$$\sqrt{n}\rho_H\left(\frac{X_1 + \cdots + X_n}{n}, \mathbf{E}X\right) \xrightarrow{d} \sup\left\{ \left|\zeta(u)\right| : u \in \mathbb{S}^{d-1} \right\} \text{ as } n \to \infty,$$

where $\left\{\zeta(u),\ u \in \mathbb{S}^{d-1}\right\}$ is a centered sample continuous Gaussian random function on $\mathbb{S}^{d-1}$ with covariance $\mathbf{E}[\zeta(u)\zeta(v)] = \mathbf{E}[h_X(u)h_X(v)] - \mathbf{E}[h_X(u)]\mathbf{E}[h_X(v)]$.
A proof is given in Molchanov (2005, theorem 3.2.1).

3. APPLICATIONS TO IDENTIFICATION ANALYSIS

Identification analysis entails the study of what can be learned about a parameter of interest, given the available data and maintained modeling assumptions. Within the partial identification paradigm, the goal is to characterize the sharp identification region, denoted Θ_I in what follows. This region exhausts all the available information, given the sampling process and the maintained modeling assumptions. Although it sometimes is easy to characterize Θ_I, there exist many important problems in which a tractable characterization is difficult to obtain. It may be particularly difficult to prove sharpness, that is, to show that a conjectured region contains exactly the feasible parameter values and no others. Proving sharpness is important. If a conjectured region is not sharp, then some parameter values in it are actually inconsistent with the sampling process and the maintained assumptions. Hence, they cannot have generated the observed data. Failure to eliminate such values weakens the model's ability to make useful predictions. It also weakens the researcher's ability to achieve point identification when it attains, as well as to test for model misspecification. This is true both in the context of structural analysis and in the context of reduced form analysis.

3.1. Sharp Identification Regions

Tractable characterizations of sharp identification regions have been provided in several contexts using standard tools of probability theory (see, e.g., Manski 1989, 2003; Manski & Tamer 2002; Molinari 2008). Beresteanu et al. (2011) show how to apply random set theory to yield a unified method for characterizing Θ_I, including applications in some important settings for which other approaches are less tractable. Their approach rests on the fact that in many partially identified models, the information in the data and assumptions can be expressed as the requirement that either (*a*) a random vector belongs to a random set with probability one, or (*b*) the conditional expectation of a random vector belongs to the conditional Aumann expectation of a random set almost surely with respect to the restriction of $\mathbf{P}$ to the conditioning σ-algebra. This immediately allows for characterizations of the elements of Θ_I through Artstein's inequalities and through the support function dominance condition, respectively. We illustrate these ideas using two simple examples.

Example 7 (best linear prediction with interval outcomes and covariates): Suppose the researcher is interested in the best linear prediction of y given x but observes only random intervals $Y = [y_L, y_U]$ and $X = [x_L, x_U]$ such that $\mathbf{P}\{y_L \leq y \leq y_U, x_L \leq x \leq x_U\} = 1$. Earlier on, Horowitz et al. (2003) studied this problem and provided a characterization of the sharp identification region of each component of the vector θ. The computational complexity of the problem in their formulation, however, grows with the number of points in the support of the outcome and covariate variables and becomes essentially unfeasible if these variables are continuous, unless one discretizes their support quite coarsely. We show here that the random set approach yields a characterization of Θ_I that remains computationally feasible regardless of the support of outcome and covariate variables.

Suppose X and Y are integrably bounded. Then one can obtain Θ_I as the collection of θ's such that there are selections $(\tilde{x}, \tilde{y}) \in \mathrm{sel}(X \times Y)$ and associated prediction errors $\varepsilon(\theta) = \tilde{y} - \theta_1 - \theta_2 \tilde{x}$, satisfying $\mathbf{E}\varepsilon(\theta) = 0$ and $\mathbf{E}\big(\varepsilon(\theta)\tilde{x}\big) = 0$. Hence, we build the set

$$Q_\theta = \left\{ q = \begin{pmatrix} \tilde{y} - \theta_1 - \theta_2 \tilde{x} \\ (\tilde{y} - \theta_1 - \theta_2 \tilde{x})\tilde{x} \end{pmatrix} : (\tilde{x}, \tilde{y}) \in \mathrm{sel}(X \times Y) \right\}.$$

We remark that Q_θ is not necessarily convex.

For a given θ, we can have a mean-zero prediction error uncorrelated with its associated selection $\tilde{x}$ if and only if the zero vector belongs to $\mathbf{E}Q_\theta$. Convexity of $\mathbf{E}Q_\theta$ and Equation 3 yield

$$0 \in \mathbf{E}Q_\theta \Leftrightarrow \langle 0, u\rangle \leq h_{\mathbf{E}Q_\theta}(u) \quad \forall u \in \mathbb{B}^d.$$

Using Theorem 2, we obtain

$$\Theta_I = \left\{\theta : 0 \leq \mathbf{E}\big(h_{Q_\theta}(u)\big) \quad \forall u \in \mathbb{B}^d\right\} = \left\{\theta : \max_{u \in \mathbb{B}^d}\big(-\mathbf{E}\big(h_{Q_\theta}(u)\big) = 0\right\}, \tag{4}$$

where

$$h_{Q_\theta}(u) = \max_{\tilde{y} \in Y, \tilde{x} \in X}\left[u_1\big(\tilde{y} - \theta_1 - \theta_2\tilde{x}\big) + u_2\big(\tilde{y}\tilde{x} - \theta_1\tilde{x} - \theta_2\tilde{x}^2\big)\right]$$

is an easy-to-calculate continuous-valued convex sublinear function of u, regardless of whether the variables involved are continuous or discrete.

The optimization problem in Equation 4 showing whether $\theta \in \Theta_I$ is a convex program hence is easy to solve [see, e.g., the CVX software by Grant & Boyd (2010)]. We note, however, that the set Θ_I itself is not necessarily convex. One then has to scan the parameter space to trace out Θ_I. Ciliberto & Tamer (2009) and Bar & Molinari (2013) propose methods to conduct this task. Projections of Θ_I on each of its components can be obtained using the support function of this set, as shown in Kaido et al. (2013).

Example 8 (entry game): Consider the setup in Example 2. Assume we observe data that identify $\mathbf{P}_y$, the multinomial distribution of outcomes of the game, and that the distribution of ε is known up to a finite-dimensional parameter vector that is part of θ. Earlier on, Tamer (2003), Berry & Tamer (2007), and Ciliberto & Tamer (2009) studied this problem and provided an abstract characterization of Θ_I based on augmenting the model with an unrestricted selection mechanism that picks the equilibrium played in the region of multiplicity. The selection mechanism is a rather general random function that Beresteanu et al. (2011) later showed builds all possible selections of the random set of equilibria. Because the selection mechanism may constitute an infinite-dimensional nuisance parameter, dealing with it directly creates great difficulties for the computation of Θ_I and for inference. Random set theory yields a complementary approach through which one can characterize Θ_I avoiding altogether the need to deal with the selection mechanism. The resulting characterization is computationally tractable, can be directly linked to existing inference methods (e.g., Andrews & Shi 2013), and is in the spirit of the earlier literature in partial identification that provides tractable characterizations of sharp identification regions without making any reference to the selection mechanism (see, e.g., Manski & Tamer 2002, Manski 2003). To build intuition for how this characterization is possible, we recall that Theorem 1 (Artstein's inequalities) and Theorem 2 (the Aumann expectation and support function) allow us to characterize the distribution and the expectation of each selection of a random set, without having to build such selections directly.

In our simple example, if the model is correctly specified and the observed outcomes result from pure-strategy Nash play, then a candidate θ can have generated the observed outcomes if and only if $y \in \mathrm{sel}(Y_\theta)$. An immediate application of Artstein's theorem yields

$$\Theta_I = \left\{\theta : \mathbf{P}\{y \in K\} \leq T_{Y_\theta}(K), K \subset \mathbb{K}\right\},$$

which allows one to verify whether $\theta \in \Theta_I$ by checking a finite number of moment inequalities—specifically $2^n - 2$, with n the cardinality of $\mathbb{K}$. This can potentially be a large number, but in Section 3.2 below, we show that econometric applications of random set theory similar in spirit to, but much more complex than, the example considered here motivated econometricians to find effective ways to substantially reduce the number of test sets K over which to check the dominance condition.

Beresteanu et al. (2011) show that Θ_I can be equivalently characterized in terms of an Aumann expectation and support function, observing that if the model is correctly specified, the multinomial distribution $\mathbf{P}_y$ observed in the data should belong to the collection of multinomial distributions associated with each selection of Y_θ. Recalling that the probability mass function of a discrete random variable is equal to the expectation of properly defined indicator functions, one can express the collection of multinomial distributions associated with each selection of Y_θ as an Aumann expectation. Specifically, define the set

$$Q_\theta = \left\{ q : q = \begin{pmatrix} (1-\tilde{y}_1)(1-\tilde{y}_2) \\ \tilde{y}_1(1-\tilde{y}_2) \\ (1-\tilde{y}_1)\tilde{y}_2 \\ \tilde{y}_1\tilde{y}_2 \end{pmatrix}, \tilde{y} \in \mathrm{sel}(Y_\theta) \right\}.$$

Then one can equivalently write

$$\begin{aligned} \Theta_I &= \left\{\theta : \mathbf{P}_y \in \mathbf{E}(Q_\theta)\right\} \\ &= \left\{\theta : \langle \mathbf{P}_y, u\rangle \leq h_{\mathbf{E}(Q_\theta)}(u), \forall u \in \mathbb{B}^d\right\} \\ &= \left\{\theta : \langle \mathbf{P}_y, u\rangle \leq \mathbf{E}\big(h_{Q_\theta}(u)\big), \forall u \in \mathbb{B}^d\right\} \\ &= \left\{\theta : \max_{u \in \mathbb{B}^d} \langle \mathbf{P}_y, u\rangle - \mathbf{E}\big(h_{Q_\theta}(u)\big) = 0\right\}, \end{aligned}$$

where the second line follows from Equation 3, the third line follows from Theorem 2, and the last line is an algebraic manipulation. The maximization problem in it is a convex optimization problem, and solving it is computationally easy. For example, Boyd & Vandenberghe (2004, p. 8) write, "We can easily solve [convex] problems with hundreds of variables and thousands of constraints on a current desktop computer, in at most a few tens of seconds."

The characterization based on the Aumann expectation also applies easily to situations in which outcomes of the game result from mixed-strategy Nash play or from other solution concepts, by replacing the set Q_θ with one collecting the multinomial distributions over outcomes of the game associated with each equilibrium mixed strategy. However, there is no result to date formally establishing a characterization for these models based on Artstein's inequalities.

Beresteanu et al. (2011) establish the validity of their Aumann expectation–based approach for a general class of econometric models, which they call "models with convex moment predictions." A detailed discussion of these models goes beyond the scope of this review; our two preceding examples, however, illustrate the key features of the random set approach to obtaining tractable characterizations of sharp identification regions.

In important complementary work, Galichon & Henry (2011) use the characterization of Θ_I based on Artstein's inequalities to study finite games of complete information with multiple pure-strategy Nash equilibria. They show that further computational simplifications can be obtained, by bringing to bear different mathematical tools from optimal transportation theory. A discussion of this theory is beyond the scope of this review.

3.2. Core-Determining Classes

Artstein's inequalities (Equation 1) characterize distributions of selections by solving a potentially large system of inequalities indexed by all compact subsets of the carrier space. As discussed in the previous section, however, econometric applications of random set theory call for computationally tractable characterizations of Θ_I. This motivated the study of a reduced family of test sets that still suffices to check for the selectionability of a distribution. Such a reduced family of sets was formally defined by Galichon & Henry (2006, 2011), who then implement it in the context of incomplete models that satisfy a monotonicity requirement. Here we present a definition that makes use of the containment functional; a similar definition can be given using the capacity functional.

> **Definition 6:** A family of compact sets $\mathcal{M}$ is said to be a core-determining class for a random closed set X if any probability measure μ satisfying the inequalities
>
> $$\mu(F) \geq C(F) = \mathbf{P}\{X \subset F\} \tag{5}$$
>
> for all $F \in \mathcal{M}$ is the distribution of a selection of X, so Equation 5 holds for all closed sets F.

It is easy to show that a core-determining class $\mathcal{M}$ is also distribution determining; in other words, the values of the containment functional on $\mathcal{M}$ uniquely determine the distribution of the random closed set X (distribution-determining classes in random set theory correspond to a similar concept for random variables in probability theory). However, distribution-determining classes are not necessarily core determining.

A rather easy and general core-determining class is obtained as a subfamily of all compact sets that is dense in a certain sense. For instance, in the Euclidean space, it suffices to consider compact sets obtained as finite unions of closed balls with rational centers and radii.

For a further reduction, one should impose additional restrictions on the family of realizations of X. Assume that X is almost surely convex. It is known that the containment functional $C_X(F)$, $F \in \mathcal{F}$, uniquely determines the distribution of X. Because a convex set X fits inside a convex set F, it would be natural to expect that probabilities of the type $C_X(F) = \mathbf{P}\{X \subset F\}$ for all convex closed sets F uniquely determine the distribution of X. This is, however, only the case if X is almost surely compact (see Molchanov 2005, theorem 1.7.8). Even in this case, however, the family of all convex compact sets is not a core-determining class when the random sets are of dimension greater than 1.

In some cases (most importantly, for random sets being intervals on the line), it is useful to note that $X \subset F$ if and only if $X \subset F_X$, where

$$F_X = \underset{\omega \in \Omega', X(\omega) \subset F}{\cup} X(\omega)$$

for any set Ω' of full probability. Thus, μ is the distribution of a selection of X if and only if $\mu(F_X) \geq \mathbf{P}\{X \subset F_X\}$ for all closed sets F.

Example 9 (random interval): Let $Y = [y_L, y_U]$ be a bounded random interval on the real line. In this case, it is useful to characterize selections by the inequalities in Equation 2 involving the containment functional of Y. Then μ is the distribution of a selection of Y if and only if

$$\mu([a,b]) \geq \mathbf{P}\{Y \subset [a,b]\} = \mathbf{P}\{a \leq y_L, y_U \leq b\}$$

for all segments $[a,b] \subset \mathbb{R}$.

Sometimes it is possible to partition the whole space of elementary events Ω into several subsets such that the values of X on ω's from disjoint subsets are disjoint.

Theorem 5: Consider a partition of Ω into sets Ω_i, $1 \leq i \leq N$, of positive probability, where N may be infinite. Let $\mathbb{K}_i = \cup\{X(\omega) : \omega \in \Omega_i\}$ denote the range of $X(\omega)$ for $\omega \in \Omega_i$. Assume that $\mathbb{K}_i$, $i \geq 1$, are disjoint. Then it suffices to check Equation 1 only for all K such that there is $i \in \{1, \ldots, N\}$ for which $K \subset \mathbb{K}_i$.

A proof is provided in Molchanov & Molinari (2014).

This result may yield a significantly reduced core-determining class.

Example 10 (entry game): Consider the setup in Example 2. We show above that Artstein's theorem yields

$$\Theta_I = \left\{\theta : \mathbf{P}\{y \in K\} \leq T_{Y_\theta}(K), K \subset \mathbb{K}\right\}.$$

This amounts to $2^n - 2$ inequalities to check, with n the cardinality of $\mathbb{K}$, in this case 4. An application of Theorem 5, however, shows that it suffices to check eight inequalities involving singleton sets $K = \{a\}$:

$$\Theta_I = \left\{\theta : C_{Y_\theta}(\{a\}) \leq \mathbf{P}\{y = \{a\}\} \leq T_{Y_\theta}(\{a\}), a \in \mathbb{K}\right\}.$$

Galichon & Henry (2011) propose the use of a matching algorithm to check that a probability measure is selectionable, using tools from optimal transportation theory. Indeed, a random vector x is a selection of X if and only if it is possible to match values $x(\omega)$ for $\omega \in \Omega$ to the values of $X(\omega)$ so that $x(\omega) \in X(\omega)$. This yields an alternative algorithm to check the selectionability and also makes it possible to quantify how far a random vector is from the family of selections.

3.3. Random Sets in the Space of Unobservables

In Example 7, for the interval data case, we encounter a random closed set defined in the space of unobservables, the prediction errors. Random closed sets defined in such space can be extremely useful for incorporating restrictions on the unobservables in the analysis, as illustrated by Chesher and Rosen in a series of papers (e.g., Chesher & Rosen 2012, Chesher et al. 2013). Here we illustrate their approach through the entry game example.

Example 11 (entry game): Consider again the two-player entry game in Example 2. So far we have addressed the identification problem in this model by defining the random closed set $Y_\theta(\varepsilon)$ of pure-strategy Nash equilibria associated with a given realization of $\varepsilon = (\varepsilon_1, \varepsilon_2)$. The random set Y_θ can be viewed as a set-valued function of ε. The inverse of this function is defined as (see Aubin & Frankowska 1990)

$$\overline{Y}_\theta(y) = \{\varepsilon : y \in Y_\theta(\varepsilon)\}.$$

If y is a random element in $\mathbb{K}$, then $\overline{Y}_\theta$ is a random closed set in the space of unobservables. Then

$$y \in \operatorname{sel}(Y_\theta(\varepsilon)) \Leftrightarrow \varepsilon \in \operatorname{sel}(\overline{Y}_\theta(y)),$$

so that using Artstein's inequalities (Equation 2), we find that a candidate distribution for ε is the distribution of a selection of $\overline{Y}_\theta(y)$ if and only if

$$\mathbf{P}\{\varepsilon \in F\} \geq \mathbf{P}\{\overline{Y}_\theta(y) \subset F\}$$

for all closed sets F in the plane, which is the realization space for ε. However, Chesher & Rosen (2012) show that this family of test sets can be considerably reduced to being equivalent to the case in which one works with Y_θ, by observing that the realizations of $\overline{Y}_\theta(y)$ associated with the four realizations of $y \in \mathbb{K}$ are four rectangles (see **Figure 1**). Hence, one can construct the core-determining class in the following steps. (a) Let F be a proper subset of one of the four rectangles; then $\mathbf{P}\{\overline{Y}_\theta(y) \subset F\} = 0$, and the inequality is automatically satisfied. (b) Take the collection of sets F that contain one of the four rectangles but not more. Then it suffices to check the inequalities for the four sets F that equal (the closure of) each of the rectangles; this is because a larger set F' in this family yields the same value for the containment functional as F. (c) Take the collection of sets F that contain two of the four rectangles but not more. A similar reasoning allows one to check the inequalities only on the five sets F that equal (the closure of) unions of two of the rectangles. Observing that the realizations $\overline{Y}_\theta(0, 0)$ and $\overline{Y}_\theta(1, 1)$ are disjoint, one obtains that the containment functional is additive on sets $F = F_1 \cup F_2$ such that $\overline{Y}_\theta(0, 0) \subset F_1$ and $\overline{Y}_\theta(1, 1) \subset F_2$; therefore, inequalities involving this set are redundant. (d) Finally, the collection of sets F that contain three of the four rectangles can similarly be reduced to (the closure of) unions of three of the rectangles and therefore yield redundant inequalities.

As this example makes plain, one can often work with random sets defined either in the space of observables or in the space of unobservables. It is then natural to ask which might be more advantageous in practice. We believe the answer depends on the modeling assumptions. As a rule of thumb, if the modeling assumptions are either stochastic or shape restrictions on the observables, it is often most useful to work with random sets defined in the space of observables. If the modeling assumptions are either stochastic or shape restrictions in the space of unobservables, it is often most useful to work with random sets defined in the space of unobservables. Suppose, for example, within the two-player entry game discussed above, that one observes variable v along with y. Then if the model is correctly specified, one has that $(y, v) \in \operatorname{sel}(Y_\theta, v)$. Impose the exclusion restriction that y is independent of v. Notice that the capacity and containment functional of Y_θ may depend on v. Then applying Artstein's inequalities, one immediately gets (see Molchanov & Molinari 2014)

$$\Theta_I = \left\{\theta : C_{Y_\theta|v}\left(\{a\}\right) \leq \mathbf{P}\{y = \{a\}\} \leq T_{Y_\theta|v}\left(\{a\}\right), a \in \mathbb{K}, v - \text{a.s.}\right\}.$$

Alternatively, suppose the exclusion restriction is between an instrumental variable v and the unobservable ε. If the model is correctly specified, $(\varepsilon, v) \in \operatorname{sel}\left(\overline{Y}_\theta, v\right)$, and a similar reasoning as before yields

$$\Theta_I = \left\{\theta : \mathbf{P}\{\varepsilon \in F\} \geq \mathbf{P}\left\{\overline{Y}_\theta(y) \subset F | v\right\}, F \in \mathcal{M}, v - \text{a.s.}\right\},$$

where $\mathcal{M}$ is the core-determining class obtained above. For other examples, readers are referred to Beresteanu et al. (2012) and Chesher & Rosen (2012).

4. APPLICATIONS TO INFERENCE

Identification arguments are always at the population level. That is, they presume that identified features of the model can be learned with certainty from observation of the entire population. However, in practice, such features need to be estimated from a finite sample. When a model is partially identified, statistical inference is particularly delicate to conduct. This is because the identified feature of the model is a set rather than a point. The shape and size of a properly defined set estimator change with sample size, and even the consistency of the estimator becomes harder to determine. Horowitz & Manski (2000), Manski & Tamer (2002), Imbens & Manski (2004), Chernozhukov et al. (2007), and Andrews & Soares (2010), among others, address the question of how to conduct inference in partially identified models, using tools of probability theory for random variables. A complementary approach is built on elements of random set theory. The method offers a unified approach to inference for level sets and convex identified sets based on Wald-type test statistics for the Hausdorff distance. The approach has been shown to be especially advantageous when Θ_I is convex, or when one is interested in inference for projections of Θ_I, because, in this case, the support function is a natural tool to obtain a functional representation of the boundary of the set, or its projections directly.

4.1. Estimation of Level Sets

The nature of partial identification problems calls for estimation of sets that appear as solutions to systems of inequalities. In the case of one inequality, consider the set $S(t) = \left\{s \in \mathbb{R}^d : f(s) \leq t\right\}$ for a lower semicontinuous real-valued function f and some $t \in \mathbb{R}$. The lower semicontinuity property of f is actually equivalent to the closedness of such level sets. If now f is replaced by its empirical estimator f_n, then $S_n(t) = \left\{s \in \mathbb{R}^d : f_n(s) \leq t\right\}$ yields the plug-in estimator of S. If f is a probability density function, then the complement of the set $S(t)$ appears in cluster analysis (see Hartigan 1975). More sophisticated estimators of $S(t)$ using the so-called excess mass method are considered in Polonik (1995). The asymptotic normality of plug-in estimators is studied in Mason & Polonik (2009), and optimal rates are obtained in Rigollet & Vert (2009).

Molchanov (1998) shows that the plug-in estimator is strongly consistent if $S(t)$ equals the closure of the set $\left\{s \in R^d : f(s) < t\right\}$. This condition is also necessary under some rather mild technical conditions. However, this condition is violated if f has a local minimum at level t. Most importantly, this is the case if t is the global minimum of the function f, and $S(t)$ is then the set arginf f of the global minimizers of the function f. This case has been thoroughly analyzed by Chernozhukov et al. (2007), who suggest estimating $S(t)$ by the set $\{s : f_n(s) \leq t + \varepsilon_n\}$, where the nonnegative correction term ε_n declines to zero at an appropriate rate.

A limit theorem obtained in Molchanov (1998) for the plug-in estimator provides a limit distribution for the normalized Hausdorff distance between $S(t)$ and $S_n(t)$, both intersected with any given compact set K. The limit theorem holds under the assumptions that the normalized difference $f_n(s) - f(s)$ satisfies a limit theorem and that f satisfies a certain smoothness condition formulated in terms of its downside continuity modulus [i.e., the infimum of $f(s') - f(s)$ for s' from a neighborhood of s].

4.2. Support Function Approach

Beresteanu & Molinari (2008) propose the use of statistics based on the Hausdorff distance to perform estimation and inference on sharp identification regions Θ_I in the space of sets, so as to replicate the common Wald approach to these tasks for point identified models in the space of vectors. In particular, they employ two Wald statistics, which measure the Hausdorff distance and the directed Hausdorff distance between the identified set and a set-valued estimator, and develop large sample and bootstrap inference procedures for these statistics.

Their results apply directly to incomplete econometric models in which Θ_I is equal to the Aumann expectation of a random set that can be constructed using random variables characterizing the model. Applying the analogy principle, Beresteanu & Molinari suggest the estimation of Θ_I through a Minkowski sample average of random sets defined using the sample observations. The support function of the convex hull of these random sets is used to represent the set estimator as a sample average of elements of a functional space so that Theorems 3 and 4 (law of large numbers and central limit theorem) are used to establish consistency of the estimator and derive its limiting distribution with respect to the Hausdorff distance. They also show that the critical values of the limiting distribution can be consistently estimated through a straightforward bootstrap procedure. Hypotheses about subsets of the population identification region are tested using the Wald statistic based on the directed Hausdorff distance, and these tests are inverted to obtain confidence sets that asymptotically cover the population identification region with a prespecified probability. Additionally, hypotheses about the entire population identification region, rather than only its subsets, are tested using the Wald statistic based on the Hausdorff distance.

We illustrate Beresteanu & Molinari's approach for the case of a best linear predictor with interval outcome data. We remark that in the case of entry games, Θ_I is not convex; therefore, any statistic based on the support function yields asymptotic statements about $\mathrm{conv}(\Theta_I)$.

Example 12 (inference for best linear predictors with interval outcomes): Suppose the researcher is interested in the best linear prediction of y given x. Let (x, y_L, y_U) be the observed variables, with $\mathbf{P}\{y_L \leq y \leq y_U\} = 1$. Then the sharp identification region of the best linear prediction parameter vector θ can be obtained defining the random segment

$$G = \left\{ \begin{pmatrix} y \\ xy \end{pmatrix} : y_L \leq y \leq y_U \right\} \subset \mathbb{R}^2$$

and collecting the least squares associated with each $(\tilde{y}, x\tilde{y}) \in \mathrm{sel}(G)$:

$$\Theta_I = \left\{ \theta \in \mathbb{R}^d : \theta = \mathbf{E}\begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}^{-1} \mathbf{E}\begin{pmatrix} \tilde{y} \\ x\tilde{y} \end{pmatrix}, (\tilde{y}, x\tilde{y}) \in \mathrm{sel}(G) \right\}, \tag{6}$$

where we have assumed that G is integrably bounded (this is the case, for example, if y_L, y_U, xy_L, and xy_U are each absolutely integrable). Given a random sample $(x_i, y_{Li}, y_{Ui})_{i=1}^n$, one can estimate Θ_I using

$$\hat{\Theta}_n = \hat{\Sigma}_n^{-1} \frac{1}{n} (G_1 + \cdots + G_n),$$

where $\hat{\Sigma}_n$ is a consistent estimator of the matrix inside Equation 6. Using Theorem 3, Beresteanu & Molinari (2008) establish a Slutsky-type result and under mild regularity conditions show that

$$\rho_H\left(\hat{\Theta}_n, \Theta_I\right) = \mathcal{O}_p\left(n^{-1/2}\right).$$

To show that the support function process converges to a Gaussian process, Beresteanu & Molinari need to assume that all x variables have a continuous distribution. This assures that the set Θ_I does not have flat faces, which in turn guarantees that $h_{\Theta_I}(u)$ is differentiable in u. Therefore, a functional delta method can be employed to show, under additional mild regularity conditions, that

$$\sqrt{n}\rho_H\left(\hat{\Theta}_n, \Theta_I\right) \xrightarrow{d} \sup_{u \in \mathbb{S}^{d-1}} \|z(u)\|,$$

$$\sqrt{n}\mathbf{d}_H\left(\Theta_I, \hat{\Theta}_n\right) \xrightarrow{d} \sup_{u \in \mathbb{S}^{d-1}} \left(-z(u)\right)_+,$$

where $z(u)$ is a linear function of a centered sample continuous Gaussian process. By comparison, in the presence of flat faces in Θ_I, Bontemps et al. (2012) show that the support function process converges to the sum of a Gaussian process and a countable point process that takes nonzero values at directions u orthogonal to the flat faces of Θ_I.

A simple nonparametric bootstrap procedure that resamples from the empirical distribution of $(x_i, y_{Li}, y_{Ui})_{i=1}^n$ consistently estimates the quantiles of the limiting distributions of these Wald statistics. Hence, one can test hypotheses of the form $\mathfrak{H}_0 : \Theta_I = \Theta_0$ versus $\mathfrak{H}_A : \Theta_I \neq \Theta_0$ using the statistic based on the Hausdorff distance and hypotheses of the form $\mathfrak{H}_0' : \Theta_0 \subseteq \Theta_I$ versus $\mathfrak{H}_A' : \Theta_0 \nsubseteq \Theta_I$ using the statistic based on the directed Hausdorff distance. Inverting these tests yields confidence collections that are unions of sets that cannot be rejected as equal to either Θ_I or subsets of Θ_I. Estimation and inference can be implemented using standard statistical packages, including STATA (see **http://economics.cornell.edu/fmolinari/#Stata_SetBLP**).

Bontemps et al. (2012) extend these results in important directions by allowing for incomplete linear moment restrictions, in which the number of restrictions exceeds the number of parameters to be estimated, and extend the familiar Sargan test for overidentifying restrictions to partially identified models. When the number of restrictions equals the number of parameters to be estimated, the authors propose a support function–based statistic to test hypotheses about each vector $\theta \in \Theta_I$ and invert this statistic to obtain confidence sets that asymptotically cover each element of Θ_I with a prespecified probability.

Chandrasekhar et al. (2012) significantly extend the applicability of Beresteanu & Molinari's (2008) approach to cover the best linear approximation of any function $f(x)$ that is known to lie within two identified bounding functions. The lower and upper functions defining the band are allowed to be any functions, including ones carrying an index, and can be estimated parametrically or nonparametrically. Because the intervals defining the outcome variable [i.e., the extreme points of the band on $f(x)$] can be estimated nonparametrically in a first stage, Chandrasekhar et al. develop a new limit theory for the support function process and prove that it approximately converges to a Gaussian process and that the Bayesian bootstrap can be applied for inference. They also propose a simple data-jittering procedure, in which a continuously distributed error with arbitrarily small but positive variance is added to each discrete variable in x, eliminating flat faces in Θ_I. Hence, they obtain valid, albeit arbitrarily conservative, inference without ruling out discrete covariates. In a study of inference for sets defined by one smooth nonlinear inequality, Chernozhukov et al. (2012) show that the (directed) Hausdorff statistic can be weighted to enforce either exact or first-order equivariance to transformations of parameters.

4.3. Duality Between the Level Set Approach and the Support Function Approach

Kaido (2012) further enlarges the domain of applicability of the support function approach by establishing a duality between level set estimators based on convex criterion functions and the support function of the level set estimators. This allows one to use Hausdorff-based statistics and the support function approach not only when Θ_I is the Aumann expectation of a properly defined random closed set, but also when such a representation is not readily available.

Kaido considers an identification region and its corresponding level set estimator given, respectively, by

$$\Theta_I = \left\{\theta \in \Theta : f(\theta) = 0\right\},$$
$$\hat{\Theta}_n(t) = \left\{\theta \in \Theta : a_n f_n(\theta) \leq t\right\},$$

where Θ is a convex subset of $\mathbb{R}^d$, f is a convex, lower semicontinuous criterion function with values in $\mathbb{R}_+$ and infimum at zero, f and f_n satisfy the additional regularity conditions set forth in Chernozhukov et al. (2007), a_n is a growing sequence, and $t \geq 0$ is properly chosen. Under these assumptions, Θ_I is convex, and Kaido (2012, lemma 3.1) establishes that

$$h_{\hat{\Theta}_n(t)}(u) < b \Leftrightarrow \inf_{\theta \in K_{b,u} \cap \Theta} a_n f_n(\theta) > t,$$

where $K_{b,u} = \left\{\theta \in \mathbb{R}^d : \langle u, \theta \rangle \geq b\right\}$.

Using this result, Kaido shows how to relate the normalized support function process $\mathcal{Z}_n(u,t) = a_n\left(h_{\hat{\Theta}_n(t)}(u) - h_{\Theta_I}(u)\right)$ to a localized version of the criterion function f_n to obtain its asymptotic distribution using the notion of weak epiconvergence (see Molchanov 2005, section 5.3). An application of Hörmander's embedding theorem then yields the asymptotic distribution of test statistics based on the Hausdorff distance.

Kaido et al. (2013) show that Kaido's (2012) approach can be extended to conduct inference on projections of Θ_I even when this set is nonconvex and the identified set is estimated using a level set estimator under the assumptions of Chernozhukov et al. (2007). Their method is based on the simple observation that the projections of Θ_I are equal to the projections of $\mathrm{conv}(\Theta_I)$, and as such, when projections are the object of interest, no information is lost owing to the convexification effect of the support function approach.

4.4. Efficiency of the Support Function Approach

Kaido & Santos (2014) develop a theory of efficiency for estimation of partially identified models defined by a finite number of convex moment inequalities of the form $\mathbf{E}(m_j(x;\theta)) \leq 0, j = 1, \ldots J$, which are smooth as functionals of the distribution of the data. The functions $\theta \mapsto m_j \mathbf{E}\left(m_j(x;\theta)\right)$ are assumed to be convex so that $\Theta_I = \{\theta : \mathbf{E}(m_j(x;\theta)) \leq 0, j = 1, \ldots J\}$ is convex and can be represented through its support function. Using the classic results in Bickel et al. (1993), Kaido & Santos show that under suitable regularity conditions, the support function admits for $\sqrt{n}$-consistent regular estimation. The assumptions rule out, in particular, (*a*) flat faces in Θ_I that depend on parameters to be estimated, (*b*) more binding moment inequalities than parameters to be estimated at any boundary point of Θ_I, and (*c*) sets Θ_I with empty interiors. Using the convolution theorem, they establish that any regular estimator of the support function must converge in distribution to the sum of a mean zero Gaussian process $\mathbb{G}_0$ and an independent noise process Δ_0.

Using the same reasoning as in the classical case, Kaido & Santos (2014) call a support function estimator semiparametrically efficient if it is regular and its asymptotic distribution equals that of $\mathbb{G}_0$. Hence, they obtain a semiparametric efficiency bound for regular estimators of the support function by deriving the covariance kernel of $\mathbb{G}_0$. Then they show that a simple plug-in estimator based on the support function of the set of parameters satisfying the sample analog of the moment inequalities attains this bound.

The semiparametrically efficient estimator of the support function is used to construct estimates of the corresponding identified set that minimize a wide class of asymptotic loss functions based on the Hausdorff distance. To estimate critical values of the limiting distribution of test statistics based on the Hausdorff distance, Kaido & Santos (2014) propose a score-multiplier bootstrap that does not require the support function to be recomputed for each resample of the data, which is especially computationally attractive. In addition to convex moment inequality models, Kaido & Santos's (2014) results imply that the estimator for the best linear predictor with interval outcome data in Example 12 is asymptotically efficient.

5. CONCLUSIONS

Although its initial development was in part motivated by questions of general equilibrium analysis and decision theory, random set theory had not been introduced in econometrics until recently. The new surge of interest in applications of random set theory to econometrics has been motivated by partially identified models, in which the identified object is a set rather than a singleton. Researchers interested in partially identified models need to provide tractable characterizations of sharp identification regions and need to develop methodologies to estimate sets, test hypotheses about (subsets of) the identification regions, and build confidence sets that cover them with a prespecified asymptotic probability.

Each of these tasks may be simplified by the use of random set theory, and many results can be developed under a unified framework. This is because random set theory distills elements of topology, convex geometry, and probability theory to directly provide a mathematical framework designed to analyze random elements whose realizations are sets. The resulting tools have proven especially useful for inference when the econometric model yields a convex sharp identification region, and for identification analysis when the informational content of the econometric model is equivalent to the statement that (the conditional expectation of) an (un)observable variable almost surely belongs to (the conditional Aumann expectation of) a random set.

This article has attempted to introduce the basic elements of random set theory that have proven useful to date in econometrics and to summarize the main applications of the theory within this literature. The hope is that this review can stimulate and simplify further applications of random set theory in econometrics.

The random set approach to partial identification may complement a more traditional approach based on laws of large numbers and central limit theorems for random vectors, which continues to be productively applied in the field. We do not review results based on these methods, but refer readers to Tamer (2010), and references therein, for a survey of the partial identification literature.

We also do not summarize the important literature in decision theory that employs elements of random set theory, most notably nonadditive measures and Choquet integrals. We refer interested readers to Gilboa (2004) for a thorough treatment of these topics.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Charles Manski for detailed comments that substantially improved this review. F.M. gratefully acknowledges financial support from NSF grant SES-0922330. I.M. was supported by Swiss National Foundation grants 200021-126503 and 200021-137527.

LITERATURE CITED

Andrews DWK, Shi X. 2013. Inference based on conditional moment inequalities. *Econometrica* 81:609–66

Andrews DWK, Soares G. 2010. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78:119–57

Artstein Z. 1983. Distributions of random sets and random selections. *Isr. J. Math.* 46:313–24

Artstein Z, Vitale RA. 1975. A strong law of large numbers for random compact sets. *Ann. Probab.* 3:879–82

Aubin J-P, Frankowska H. 1990. *Set-Valued Analysis*. Boston: Birkhäuser

Aumann RJ. 1965. Integrals of set-valued functions. *J. Math. Anal. Appl.* 12:1–12

Bar HY, Molinari F. 2013. *Computation of sets via data augmentation and support vector machines*. Unpublished manuscript, Cornell Univ., Ithaca, NY

Beresteanu A, Molchanov I, Molinari F. 2011. Sharp identification regions in models with convex moment predictions. *Econometrica* 79:1785–821

Beresteanu A, Molchanov I, Molinari F. 2012. Partial identification using random set theory. *J. Econom.* 166:17–32. Errata. **http://econweb.arts.cornell.edu/fmolinari/NOTE_BMM2012_v3.pdf**

Beresteanu A, Molinari F. 2008. Asymptotic properties for a class of partially identified models. *Econometrica* 76:763–814

Berry ST, Tamer E. 2007. Identification in models of oligopoly entry. In *Advances in Economics and Econometrics: Theory and Applications*, *Ninth World Congress*, Vol. II, ed. R Blundell, WK Newey, T Persson, pp. 46–85. Cambridge, UK: Cambridge Univ. Press

Bickel PJ, Klaassen CA, Ritov Y, Wellner JA. 1993. *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer

Bontemps C, Magnac T, Maurin E. 2012. Set identified linear models. *Econometrica* 80:1129–55

Boyd S, Vandenberghe L. 2004. *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press

Chandrasekhar A, Chernozhukov V, Molinari F, Schrimpf P. 2012. *Inference for best linear approximations to set identified functions*. Work. Pap. CWP 43/12, CeMMAP, London

Chernozhukov V, Hong H, Tamer E. 2007. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75:1243–84

Chernozhukov V, Kocatulum E, Menzel K. 2012. *Inference on sets in finance*. Work. Pap. CWP 46/12, CeMMAP, London

Chesher A, Rosen AM. 2012. *Simultaneous equations models for discrete outcomes: coherence, completeness, and identification*. Work. Pap. CWP21/12, CeMMAP, London

Chesher A, Rosen AM, Smolinski K. 2013. An instrumental variable model of multiple discrete choice. *Quant. Econ.* 4:157–96

Choquet G. 1953/1954. Theory of capacities. *Ann. Inst. Fourier* 5:131–295

Ciliberto F, Tamer E. 2009. Market structure and multiple equilibria in airline markets. *Econometrica* 77:1791–828

Debreu G. 1967. Integration of correspondences. *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, Vol. 2, pp. 351–72. Berkeley: Univ. Calif. Press

Galichon A, Henry M. 2006. *Inference in incomplete models*. Unpublished manuscript, Columbia Univ., New York

Galichon A, Henry M. 2011. Set identification in models with multiple equilibria. *Rev. Econ. Stud.* 78:1264–98

Gilboa I. 2004. *Uncertainty in Economic Theory: Essays in Honor of David Schmeidler's 65th Birthday*. London: Routledge

Grant M, Boyd S. 2010. *CVX: Matlab software for disciplined convex programming, version 1.21*. **http://cvxr.com/cvx**

Hartigan JA. 1975. *Clustering Algorithms*. New York: Wiley

Hörmander L. 1954. Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. *Ark. Mat.* 3:181–86

Horowitz JL, Manski CF. 2000. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Stat. Assoc.* 95:77–84

Horowitz JL, Manski CF, Ponomareva M, Stoye J. 2003. Computation of bounds on population parameters when the data are incomplete. *Reliab. Comput.* 9:419–40

Imbens GW, Manski CF. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72:1845–57

Kaido H. 2012. *A dual approach to inference for partially identified econometric models*. Work. Pap., Boston Univ.

Kaido H, Molinari F, Stoye J. 2013. *Inference for projections of identified sets*. Manuscript in preparation

Kaido H, Santos A. 2014. Asymptotically efficient estimation of models defined by convex moment inequalities. *Econometrica* 82:387–413

Kolmogorov AN. 1950. *Foundations of the Theory of Probability*. New York: Chelsea

Manski CF. 1989. Anatomy of the selection problem. *J. Hum. Resour.* 24:343–60

Manski CF. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard Univ. Press

Manski CF. 2003. *Partial Identification of Probability Distributions*. New York: Springer-Verlag

Manski CF, Tamer E. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70:519–46

Mason DM, Polonik W. 2009. Asymptotic normality of plug-in level set estimates. *Ann. Stat.* 19:1108–42

Matheron G. 1975. *Random Sets and Integral Geometry*. New York: Wiley

Molchanov I. 1998. A limit theorem for solutions of inequalities. *Scand. J. Stat.* 25:235–42

Molchanov I. 2005. *Theory of Random Sets*. London: Springer

Molchanov I, Molinari F. 2014. *Random Sets in Econometrics*. In preparation

Molinari F. 2008. Partial identification of probability distributions with misclassified data. *J. Econom.* 144:81–117

Norberg T. 1992. On the existence of ordered couplings of random sets: with applications. *Isr. J. Math.* 77:241–64

Polonik W. 1995. Measuring mass concentrations and estimating density contour clusters: an excess mass approach. *Ann. Stat.* 23:855–81

Rigollet P, Vert R. 2009. Optimal rates in plug-in estimators of density level sets. *Bernoulli* 15:1154–78

Schneider R. 1993. *Convex Bodies: The Brunn-Minkowski Theory*. Cambridge, UK: Cambridge Univ. Press

Tamer E. 2003. Incomplete simultaneous discrete response model with multiple equilibria. *Rev. Econ. Stud.* 70:147–65

Tamer E. 2010. Partial identification in econometrics. *Annu. Rev. Econ.* 2:167–95